一类非光滑非凸优化问题的神经网络方法 *

喻 昕，陈昭蓉†

(广西大学 计算机与电子信息学院, 南宁 530004)

摘 要：提出了解决一类带等式与不等式约束的非光滑非凸优化问题的神经网络模型。证明了当目标函数有下界时，神经网络的解轨迹在有限时间收敛到可行域。同时，神经网络的平衡点集与优化问题的关键点集一致，且神经网络最终收敛于优化问题的关键点集。与传统基于罚函数的神经网络模型不同，提出的模型无须计算罚因子。最后，通过仿真实验验证了所提出模型的有效性。

关键词：神经网络；非凸非光滑优化；有限时间收敛

中图分类号：TP183 doi: 10.3969/j.issn.1001-3695.2018.03.0150

Neural network optimization method for class of nonconvex nonsmooth optimization problems

Yu Xin, Chen Zhaorong†

(*School Of Computer & Electronic Information Guangxi University*, *Nanning* 530004, *China*)

Abstract: This paper proposed a novel neural network to solve nonsmooth nonconvex optimization problems with equality and inequality constraints. It was proved that when the objective function has a lower bound, the neural network converges to a feasible domain in a finite time. Meanwhile, the solution trajectory of neural network converges to optimal solution set of the corresponding optimization problems, which finally converge to critical point set of optimization problems. Comparing with traditional neural network which based on penalty function, the neural network model does not need to calculate any penalty parameters. Finally, the effectiveness of the proposed model is verified by simulation experiments.

Key words: Neural network; Nonconvex nonsmooth optimization; Limited time convergence

0 引言

1986 年，Tank 和 Hopfield[1]首次介绍了模拟神经网络方法解决限制性最优化问题，并在文献[2]中得到更进一步的阐述。其核心思想是利用动力学神经网络同时模拟目标函数与约束函数，并利用神经网络的模拟与并行处理能力计算问题的最优解。文献[1]首次提出了递归神经网络解决线性规划问题，在此基础上，Kennedy 和 Chua[2]对文献[1]中的神经网络进行改进，通过构建一个有限的惩罚参数的神经网络来解决非线性规划问题。Zhang[3]在早期惩罚参数研究的基础上，提出了一种基于 Lagrange 乘子法的递归神经网络以解决非线性凸优化问题。Lagrange 神经网络模型中有两种神经元：变量神经元和 Lagrange 神经元。其中，变量神经元负责寻找目标函数的最小点以及提供解决问题的平衡点；而拉格朗日神经元负责将动态轨迹快速的引入可行域。有关 Lagrange 神经网络的详细研究请参考文献[4-6]。

在最优化问题的研究发展中，早期的神经网络模型均为光滑非线性规划问题而设计，为了解决非光滑非线性规划问题，Forti 等人[7]提出了一种 G-NPC 神经网络模型，G-NPC 是基于 NPC 的一个微分包含形式的梯度系统，其目标函数和约束函数不需要是光滑的而仅仅是正则函数即可。为了使非光滑规划更具普遍性，文献[8,9]提出了一些基于惩罚函数和次梯度的递归神经网络解决非光滑凸与非光滑非凸规划问题。神经网络的收敛行为，包括全局收敛和局部收敛。如果规划问题的目标函数的定义域域中任意的初始点的解轨迹收敛于最优解（或近似最优解），则该神经网络模型是全局收敛的，如文献[10]的神经网络模型，在满足一些条件的情况下是全局收敛；然而，如果限制条件是初始状态的解必须从可行域中选取，则其收敛是局部的。基于惩罚参数和函数的神经网络模型，其解集很在大程度上是非全局收敛的，因为此类神经网络模型，其可行域上发现的一个内部点通常是一个特殊域包含的可行域的中心，因此，为了保证收敛，初始解集必须在可行域范围内部选取，并且惩罚参数必须保证足够大。再者，惩罚函数神经网络模型的有效性依赖于确定的或非确定的惩罚参数，而确定性惩罚参数在实际应用是很难估算的。为解决这个问题，Zheng 等人[11]提出了一种新的解决非光滑凸优化的递归神经网络，文献[11]所提出

收稿日期：2018-03-03；**修回日期：**2018-04-23 **基金项目：**国家自然科学基金资助项目（61462006）

作者简介：喻昕（1973-），男，教授，博士，主要研究方向为神经网络、优化计算；陈昭蓉（1991-），女（通信作者），湖北武人，硕士研究生，主要研究方向为神经网络最优化（879890352@qq.com）.

的神经网络在解决带约束的非光滑凸规划问题的同时，不需要任何的惩罚参数。然而，文献[11]提出的神经网络的结构却异常复杂。

为了降低网络模型的复杂度，文献[12~16]提出了一些一层递归神经网络。Gou 等人[14]提出了一种解决线性等式约束的伪凸优化的一层递归神经网络，并证明了在等式约束条件下，解轨迹能在有限时间收敛到可行域中。基于惩罚函数方法，Liu 等人[15]提出了一种新的一层递归神经网络来解决线性等式约束的伪凸最优化问题。为解决线性规划问题，Liu[16]早期同样提出一种基于次梯度的神经网络，证明了该神经网络定义域内是全局收敛的。国内有关神经网络解决最优化的研究可参考文献[21~23]

传统的神经网络模型解决工程优问题大多要求求解适当的罚函数，而这个参数在某些目标函数下是难以计算的，这为网络执行计算带来困难。本文结合传统的递归神经网络模型，提出新的一种解决带等式约束和不等式约束的递归神经网络模型。该模型的优点如下：a）与传统基于罚函数的神经网络模型不同，本文提出的模型无需计算罚因子；b）许多模型的网络模型的初始点往往只能在一个有界球体内选取，而此模型的网络初始点可以任意选取；c）到目前为止，所提出的大部分模型仅能解决目标函数为凸的优化问题，而此模型可解决一类目标函数为非凸的优化问题。

1　预备知识

在本文中，考虑如下的非光滑非凸约束的最优化问题：

$$\begin{aligned}&\min\quad f(x)=f(x)\\&\text{s.t.}\ \begin{cases}g(x)\le 0\\ h(x)=0\end{cases}\end{aligned}\tag{1}$$

其中: $x=(x_1,x_2,\dots,x_n)^T\in\mathbb{R}^n$ 是变量向量，$f:\mathbb{R}^n\to\mathbb{R}$ 是非光滑凸或非凸的目标函数，其中 $f(x)$ 可以是非光滑非凸的正则函数。$g(x)=(g_1(x),g(_2x),\dots,g_m(x))^T:\mathbb{R}^n\to\mathbb{R}^m$ 是 m 维向量的不等式约束向量函数，且 $g_i(x),i=1,\dots,m$ 可以是非光滑的凸函数。$h(x)=(h(x_1),h(x_2),\dots,h_p(x))^T:\mathbb{R}^n\to\mathbb{R}^p$ 是 p 维线性等式约束向量函数。不失一般性，我们假设非线性规划(1)至少有一个最优解。定义 $S_1=\left\{\boldsymbol{x}\in R^n:\boldsymbol{g}(\boldsymbol{x})\le 0\right\}$，$S_2=\left\{\boldsymbol{x}\in R^n:\boldsymbol{h}(\boldsymbol{x})=0\right\}$。满足约束函数的可行域定义为 $S=\left\{\boldsymbol{x}\in R^n:\boldsymbol{g}(\boldsymbol{x})\le 0,\boldsymbol{h}(\boldsymbol{x})=0\right\}$，显然 $S=S_1\cap S_2$。

对于本文研究的优化问题，我们一直假定存在 $\tilde{\boldsymbol{x}}\in R^n$ 和 $\delta>0$，满足 $\tilde{\boldsymbol{x}}\in\operatorname{int}(S_1)\cap S_2$，$S_1\subseteq B(\tilde{\boldsymbol{x}},\delta)$，其中 $\operatorname{int}(S_1)$ 表示集合 S_1 的内部，$B(\tilde{\boldsymbol{x}},\delta)$ 表示以 $\tilde{\boldsymbol{x}}$ 为中心半径为 δ 的开球区域，且 $G_m=\min\limits_{1\le j\le m}-g_j(\tilde{\boldsymbol{x}})>0$。

为了便于解决问题，定义

$$G(\boldsymbol{x})=\begin{cases}0,\boldsymbol{x}\in S_1\\ \sum\limits_{j\in J^+(x)}g_j(\boldsymbol{x}),\boldsymbol{x}\notin S_1\end{cases},$$

$J^+(x)=\left\{j\in\{1,2,\cdots,m\}:g_j(x)>0\right\}$，显然，$\left\{\boldsymbol{x}:G(\boldsymbol{x})\le 0\right\}=S_1$，$\left\{\boldsymbol{x}:H(\boldsymbol{x})=0\right\}=S_2$。

定义 1[19]　如果存在正整数 k,ε，对于任意的 $x_1,x_2\in x+\varepsilon B(0,1)$ 使得 $|f(x_1)-f(x_2)|\le k(x_1-x_2)$ 成立，则称函数 $f:\mathbb{R}^n\to\mathbb{R}$ 在 $x\in\mathbb{R}^n$ 附近 Lipschitz。如果 f 在其定义域每一个点的附近都 Lipschitz，则称 f 为局部 Lipschitz。

如果 $f:\mathbb{R}^n\to\mathbb{R}$ 在 $\mathbb{R}^n$ 上是局部 Lipschitz 的，则 f 在 x 点沿方向 $v\in\mathbb{R}^n$ 的广义方向导数为：

$$f^0(x;v)=\lim_{\substack{y\to x\\ t\to 0+}}\frac{f(y+tv)-f(y)}{t}$$

如果 $f^0(x;v)$ 是有定义且有限的，f 在 x 处的次微分可以定义为

$$\partial f(x)=\{\xi\in\mathbb{R}^n:\forall v\in\mathbb{R}^n,\ f^0(x;v)\ge\langle v,\xi\rangle\}$$

引理 1　链式法则[20]。如果 $W:\mathbb{R}^n\to\mathbb{R}$ 在 $x(t)$ 处是正则的，$x(t):\mathbb{R}^n\to\mathbb{R}$ 在 t 上是可微且是局部 Lipschitz 的，则 $\dot{W}(x(t))=<\xi,x(t)>,\xi\in\partial W(x(t))\ \ t\in[0,\infty)$。

本文提出以下微分包含的神经网络动力学模型：

$$\dot{x}\in-(I-P)(\partial f(\mathbf{x})+u\partial G(x))-A^T h(Ax(t)-b)\tag{2}$$

$$\dot{u}=\begin{cases}0,\boldsymbol{x}\in S_1\\ G(x)+\varepsilon,\boldsymbol{x}\in\mathbb{R}^n/S_1\end{cases},u(0)>0$$

其中：I 是单位矩阵，$P=A^T(AA^T)^{-1}A$。

2　神经网络收敛分析

定义 2[8]　在区间 $[t_0,t_1],(t_1>t_0)$ 上，$\dot{x}(t)\in-\partial W(x(t))$ 满足初始条件 $x(t_0)=x_0$ 的一个解，是一个绝对连续函数 $x(t)$，则对于 $x(t_0)=x_0$ 和所有的 $t\in[t_0,t_1]$，有 $\dot{x}(t)\in-\partial W(x(t))$。

式(10)中的一个平衡点 $x^*\in\mathbb{R}^n$，为式(10)的一个解，也就是说，$x(t)=x^*,t\in(t_0,\infty)$。显然，当且仅当满足 $0\in-\partial W(x(t))$ 时，x^* 才为其平衡点。

引理 2[8]　对任意 $\boldsymbol{x}\in S_2\setminus S_1$，$\xi\in\partial G(x)$，有 $(x-\tilde{x})^T(I-P)\xi\ge G_m$。

引理3　神经网络模型式（2）具有局部解。

证明　由于式（2）的右边是一个非空、紧凸的上半连续的集值映射，因此神经网络（10）至少存在一个具有满足初始条件 $x(0)=x_0$ 的局部解 $x(t),\ t\in[0,T)$。

定理1　对于任意的初始点 $\boldsymbol{x}_0$，神经网络模型（2）的轨迹在有限时间内必进入等式限制区域 S_2，且停留在其中。

证明　设 $H(x)=\left\|Ax-b\right\|_1$，则

$S_2=\{x\in\mathbb{R}^n:H(x)=0\}$，且 $\partial H(x)=A^T h(Ax-b)$。根据链式法则，存在 $\gamma(t)\in\partial f(x(t))$，$\xi(t)\in\partial G(x(t))$，以及 $\eta(t)\in h(Ax(t)-b)$ 满足

$$\frac{d}{dt}H(x(t))=<A^T\eta(t),\dot{x}(t)>$$
$$=-\eta(t)^T A(I-P)(\gamma(t)+u\xi(t))-\eta(t)^T AA^T\eta(t)$$

因 $A(I-P)=A-AA^T(AA^T)^{-1}A=0$，有

$$\frac{d}{dt}H(x(t))=-\eta(t)^T AA^T\eta(t)$$
$$\le-\lambda_m(AA^T)\|\eta(t)\|^2$$

其中：$\lambda_m(AA^T)>0$ 是 AA^T 的最小特征值。

由 $x(t)\notin S_2$，根据 h 的定义知，$\|\eta(t)\|\ge1$，因此：

$$\frac{d}{dt}H(x(t))\le-\lambda_M(AA^T)<0 \qquad (3)$$

对上式两边从 0 到 t 进行积分得

$$0\le H(x(t))\le H(x(0))-\lambda_m(AA^T)t$$

显然，神经网络的轨迹 $x(t)$ 能在有限时间进入 S_2。

下面证明轨迹 $x(t)$ 一直停留在 S_2 中。采用反证法，假定 $x(t)$ 在时刻 t_1 离开 S_2，且在时间 $t\in(t_1,t_2)$，$t_1<t_2$ 内在 S_2 外。由 $H(x(t_1))=0$ 及 $\frac{d}{dt}H(x(t))\le-\lambda_m(AA^T)$，$t\in(t_1,t_2)$，可知 $H(x(t))<0$，$t\in(t_1,t_2)$，这与 $H(x(t))>0$，$x(t)\notin S_2$ 不符，定理 1 得证。

为了便于后面的证明，介绍如下假设：

假设 1　目标函数 $f(x)$ 有下界。

定理 2　若假设 1 成立，则对于任意初始点 $\mathbf{x}_0$,神经网络模型式（2）具有全局解

证明　首先证明神经网络状态向量 $x(t)$ 有界。

根据定理 1 可知，神经网络状态向量 $x(t)$ 进入 S_2 并停留其中，因此神经网络模型（2）可简化为：

$$\dot{x}\in-(I-P)(\partial f(\mathbf{x})+u\partial G(x))$$

定义能量函数：

$$W(\mathbf{x}(t))=\mu^{-1}(t)(f(x(t))-\inf_{x\in R^n}f(x))+G(x(t)) \qquad (3)$$

根据链式法则，存在 $\gamma(t)\in\partial f(x(t))$，

$\xi(t)\in\partial G(x(t))$，满足：

$$(d/dt)W(\mathbf{x}(t))=<\mu^{-1}(t)\gamma(t)+\xi(t),\dot{x}(t)>+$$
$$(f(x(t))-\inf_{x\in R^n}f(x))(d/dt)\mu^{-1}(t) \qquad (4)$$

由 $u(0)>0$ 可知对任意 $t\in[0,T)$，$\mu(t)>0$，因此 $(d/dt)\mu^{-1}(t)=-\mu^2(t)(G(x)+\varepsilon)<0$，又由 $f(x(t))-\inf_{x\in R^n}f(x)\ge0$，式(4)可转换为

$$(d/dt)W(\mathbf{x}(t))\le<\mu^{-1}(t)\gamma(t)+\xi(t),\dot{x}(t)>$$
$$=\xi(t)^T)(I-P)^T(I-P)(\gamma(t)+\mu(t)\xi(t))$$
$$-(\mu^{-1}(t)\gamma(t)^T$$
$$=\frac{\|(I-P)(\gamma(t)+\mu(t)\xi(t))\|^2}{\mu(t)} \qquad (5)$$
$$\le\quad 0$$

结合式（3）和（5）可知，对对任意 $t\in[0,T)$

$W(\mathbf{x}(0))\ge W(\mathbf{x}(t))\ge G(x(t))$

而 $G(x(t))$ 是强制的，因此也是水平集有界的，从而可知神经网络状态向量 $x(t)$ 有界。

另一方面，根据解的可扩展理论，局部解可从 $t\in[0,T)$ 扩展到 $t\in[0,+\infty)$，即神经网络模型存在全局解。

定理 3　若假设 1 成立，对于任意的初始点 $\mathbf{x}_0$，神经网络模型式（2）的轨迹在有限时间内必进入可行域 S，且停留在其中。

证明　根据定理 1 可知，神经网络状态向量 $x(t)$ 进入 S_2 并停留其中。不失一般性，我们只需证明对任意初始点 $\mathbf{x}_0\in S_2\setminus S_1$，神经网络状态向量 $x(t)$ 进入 S 并最终停留其中。

此时神经网络模型（10）可简化为

$\dot{x}\in-(I-P)(\partial f(\mathbf{x})+u\partial G(x))$

定义能量函数为

$$W_1(\mathbf{x}(t))=G(x(t)) \qquad (6)$$

显然，$\{\mathbf{x}:W_1(\mathbf{x}(t))\le0\}=S_1$。根据链式法则，对任意给定 $\mathbf{x}(t)\in S_2\setminus S_1$，存在 $\gamma(t)\in\partial f(x(t))$，$\xi(t)\in\partial G(x(t))$ 满足

$$(d/dt)W_1(\mathbf{x}(t))=<\xi(t),\dot{x}(t)>=$$
$$-\xi(t)^T(I-P)(\gamma(t)+\mu(t)\xi(t))\le$$
$$\|(I-P)\xi(t)\|\|\gamma(t)\|-\mu(t)\|(I-P)\xi(t)\|^2\le \qquad (7)$$
$$\|(I-P)\xi(t)\|(\|\gamma(t)\|-\mu(t)\|(I-P)\xi(t)\|)$$

根据定理 2 的证明过程可知，$\mathbf{x}(t)$ 有界，因此存在一个足够大的 $M>0$，使得 $\|\mathbf{x}(t)-\tilde{x}\|\le M$。由于目标函数 $f(x)$ 是局部李普西兹函数，设 l_f，l_G 分别为 $f(\mathbf{x})$，$G(\mathbf{x})$ 在 $B(\tilde{x},M)$ 上的 Lipschitz 常数，可知 $\|\gamma(t)\|\le l_f$，$\|\xi(t)\|\le l_f$。另一方面，根据引理 1 可推出，对任意给定

$\mathbf{x}(t)\in\{S_2\setminus S_1\}\cap B(\tilde{x},M)$，$\|(x(t)-\tilde{x})^T(I-P)\,\xi\|\ge G_m$，从而有

$\|(I-P)\,\xi\|\ge G_m/M>0$。

此外，对任意 $\mathbf{x}(t)\in\mathbb{R}^n$，$G(t)\ge 0$，因此 $\mu(t)$ 是正单调递增的。从而必存在一个时刻 $t'\ge 0$，使得对任意 $t>t'$

$$\mu(t)>l_f\ M/G_m \tag{8}$$

综上，对任意 $t>t'$，存在一个实数 $\varepsilon_0>0$，满足

$$(d/dt)W_1(\mathbf{x}(t))\le\ -\varepsilon_0 \tag{9}$$

对（9）两边从 t' 到 t 进行积分得

$$W_1(\mathbf{x}(t))-W_1(\mathbf{x}(t'))\le\ -\varepsilon_0\ (t-t') \tag{10}$$

而 $W_1(\mathbf{x}(t'))\ge 0$ 且有上界，结合 $\{\mathbf{x}:W_1(\mathbf{x}(t))\le 0\}=S_1$，可导出 $\mathbf{x}(t)$ 必在有限时间内进入 S_1，即进入可行域 S。

下面仍需证明，若 $\mathbf{x}(t_0)$，$t_0>t'$ 在 S 中，则当 $t>t'$，$\mathbf{x}(t)$ 将一直停留在那里。

采用反证法，假定 $x(t)$ 在时刻 $t_1'>t_0$ 离开 S，且在时间 $t\in(t_1',t_2')$，$t_1'<t_2'$ 内在 S 外。由 $W_1(x(t_1'))=0$ 及 $\frac{d}{dt}W_1\left(x(t)\right)\le-\varepsilon$，$t\in(t_1',t_2')$，可知 $W_1(x(t))<0$，$t\in(t_1',t_2')$，这与 $W_1(x(t))>0$，$x(t)\notin S_1$ 相矛盾，定理 3 得证。

定义 3　神经网络的平衡点集合 ε_σ 为

$$\varepsilon_\sigma=\left\{\begin{array}{l}x^*,\lambda^*,\mu^*\,|\,0\in\partial f(x^*)+\sum_{j=1}^{n}\lambda_j^*\partial h_j(x^*)\\+\sum_{i=1}^{m}u_i^*\partial g_i(x^*)+2c\sum_{j=1}^{n}(\partial h_i(x^*))^T h_j(x)\\+c\sum_{i}^{m}(\partial g_i(x^*))^T u_i^2 g_i(x^*)\\0=\sum_{j=1}^{n}\lambda_j^*\partial h_j(x^*),0=\sum_{i=1}^{m}u_i^*\partial g_i(x^*)\end{array}\right\}$$

定理 4　当假设 1 成立，给定任意的初始点 $\mathbf{x}_0\in\mathbb{R}^n$，神经网络轨迹 $\mathbf{x}(t)$ 满足 $\lim_{t\to+\infty}\operatorname{dist}(\mathbf{x}(t),\varepsilon_\sigma)=0$。

证明　为了便于后文的证明，首先定义能量函数：$W'(x)=f(x)$。由定理 3 可知，对于任意初始点 $x_0\in S$，$x(t)$ 会一直停留在 S 中。因为 $W'(x)$ 是正则函数，存在可测函数 $\gamma(t)\in\partial f\left(x(t)\right)$，$\xi(t)\in\partial G\left(x(t)\right)$，以及 $\eta(t)\in h\left(Ax(t)-b\right)$，有

$(d/dt)W'(x(t))=<\gamma(t),-(I-P)(\gamma(t)+u\xi(t))>$

与此同时，因为 $x(t)\in S$，$\partial G(x)=0$，也就是说 $<\xi(t),\dot{x}(t)>=0$。上式则可以改写成

$(d/dt)W'(x(t))=<\gamma(t)+u\xi(t),-(I-P)(\gamma(t)+u\xi(t))>$

而 $(I-P)^2=I-P$，$(I-P)^T=(I-P)$，所以

$$\begin{aligned}(d/dt)W'(x(t))&=-\|(I-P)(\gamma(t)+u\xi(t))\|^2\\&=-\|\dot{x}(t)\|^2\end{aligned} \tag{11}$$

根据假设 1，$f(x)$ 有下界，可知 $W'(x)$ 有界。

接下来，通过反证法证明定理 2，假设 $\lim_{t\to+\infty}\operatorname{dist}(\mathbf{x}(t_n),\varepsilon_\sigma)=\mathrm{p}$，那么存在一个序列 $\{\mathrm{t_n}\}$，当 $\lim_{n\to\infty}t_n=+\infty$ 时，有 $\lim_{t\to+\infty}\operatorname{dist}(\mathbf{x}(t_n),\varepsilon_\sigma)=\mathrm{p}$ 并且 $\overline{dist(x(t_n),\varepsilon_\delta)>p/2}$。不妨假定 $H=\mathrm{B}(\tilde{x},R)\setminus(\varepsilon_\delta+\frac{p}{4}\mathrm{B}(0,1))$，$\Upsilon(x):=-(I-P)(\partial f(x)+u\partial G(x))$。由前文的分析，可知 $\dot{x}(t)\in\Upsilon(x(t))$。

对于任意的 $x(t)\in H$，有 $0\notin\Upsilon(x(t))$。而 $\Upsilon(x(t))$ 是一个紧凸集，则在 H 上存在最小值与最大值，分别表示为

$$m_H=\min_{x\in H}\min_{\zeta\in\Upsilon(x)}\|\zeta\|>0$$

$$M_H=\max_{x\in H}\max_{\zeta\in\Upsilon(x)}\|\zeta\|>0$$

此外，对于任意时间间隔 $t\in\left[t_n,t_n+\alpha\right]$，$x(t)\in H$，有 $0\le\alpha\le p/(4M_H)$。结合式（11），可得

$$\begin{aligned}\lim_{t\to+\infty}W'(x(t))&=W'(x_0)-\int_0^{+\infty}\|\dot{x}(t)\|^2dt\\&\le W'(x_0)-\int_{\bigcup_{n\in\mathbb{N}}[t_n,t_n+\alpha]}\|\dot{x}(t)\|^2\,dt\\&\le W'(x_0)-\int_{\bigcup_{n\in\mathbb{N}}[t_n,t_n+\alpha]}m_H^2dt\\&=W'(x_0)-\sum_{n\in\mathbb{N}}\alpha m_H^2=-\infty\end{aligned}$$

这与 $W'(x(t))$ 有界矛盾，假设不成立。因此，$\lim_{t\to+\infty}\operatorname{dist}(\mathbf{x}(t),\varepsilon_\sigma)=0$，证毕。

定义 4　优化问题的关键点集合 C 为 $C=\left\{\mathbf{x}\in S:0\in\partial f(\mathbf{x})+N_S(\mathbf{x})\right\}$，其中 $N_S(\mathbf{x})$ 是 $\mathbf{x}$ 在可行域 S 上的法锥。

定理 5　当假设 1 成立，给定任意的初始点 $\mathbf{x}_0\in\mathbb{R}^n$，神经网络模型（2）的任意一个聚点都是优化问题（1）的关键点。

证明　为了便于后文的证明，首先定义能量函数：$W'(x)=f(x)$。由定理 3 可知，对于任意初始点 $x_0\in S$，$x(t)$ 会一直停留在 S 中。由网络模型（2）可知当 $x(t)\in S$，$u(t)\equiv\bar{u}$。

因 $W'(x)$ 是正则函数，存在可测函数 $\gamma(t)\in\partial f\left(x(t)\right)$，$\xi(t)\in\partial G\left(x(t)\right)$ 满足

$$(d/dt)W'(x(t)) =< \gamma(t), -(I-P)(\gamma(t)+\bar{u}\xi(t)) >$$

因 $x(t)\in S$，$\dot{G}(x)=0$，即 $<\xi(t),\dot{x}(t)>=0$。上式则可以改写成

$(d/dt)W'(x(t)) =< \gamma(t)+\bar{u}\xi(t), -(I-P)(\gamma(t)+\bar{u}\xi(t)) >$，而 $(I-P)^2=I-P$，$(I-P)^T=(I-P)$，所以

$$(d/dt)W'(x(t)) = -\|(I-P)(\gamma(t)+\bar{u}\xi(t))\|^2 = -\|\dot{x}(t)\|^2 \le 0$$

可知 $W'(x)$ 是一个非增函数。又由 S 有界可知 $W'(x(t))$ 有下界，从而可知 $\lim_{t\to+\infty}W'(x(t))$ 存在，且

$$\lim_{t\to+\infty}\|\dot{x}(t)\|=0 \quad (12)$$

由 $x(t)$ 在 S 中且 S 有界，可知存在一个聚点 $\bar{x}$ 和一个时间序列 $\{t_k\}$ 满足 $\lim_{k\to+\infty}t_k=+\infty$，$\lim_{n\to+\infty}\boldsymbol{x}(t_n)=\bar{x}$。

由式（12）可知 $\lim_{k\to+\infty}\left\|\dot{x}_{t_k}\right\|=0$，从而有

$$\lim_{k\to+\infty}-(I-P)(\gamma_{t_k}+\bar{u}\xi_{t_k})=0$$

其中 $\gamma_{t_k}\in\partial f(x_{t_k})$，$\xi_{t_k}\in\partial G(x_{t_k})$。由 $\partial f(x)$ 及 $\partial G(x)$ 是上半连续的集值映射，可知

$$\lim_{k\to+\infty}\gamma_{t_k}\in\partial f(\bar{x})$$

$$\lim_{k\to+\infty}\xi_{t_k}\in\partial G(\bar{x})$$

$$0\in h(A\bar{x}-b)$$

有 $0\in-(I-P)(\partial f(\bar{x})+\bar{u}\partial G(\bar{x}))-A^T h(A\bar{x}-b)$ （13）

可知 $(\bar{x},\bar{u})$ 是网络模型式（2）的一个平衡点。

从式（13）可知，存在 $\bar{\gamma}\in\partial f(\bar{x})$，$\bar{\xi}\in\partial G(\bar{x})$，$\bar{\eta}\in h(A\bar{x}-b)$，满足：

$$0=-(I-P)(\bar{\gamma}+\bar{u}\bar{\xi})-A^T\bar{\eta} \quad (14)$$

令 $\pi=\bar{\eta}-(AA^T)^{-1}A(\bar{\gamma}+\bar{u}\bar{\xi})$，于是式（14）可以改写成 $0=\bar{\gamma}+\bar{u}\bar{\xi}+A^T\pi$。根据法锥的定义可知 $A^T\pi\in N_{s_2}(\bar{x})$，$u\bar{\xi}\in N_{s_1}(\bar{x})$，从而有

$$0\in\partial f(\bar{x})+N_{s_1}(\bar{x})+N_{s_2}(\bar{x})$$

考虑优化问题的假定 $\mathrm{int}(S_1)\bigcap S_2\neq\varnothing$，有 $N_s(\bar{x})=N_{s_1}(\bar{x})+N_{s_2}(\bar{x})$，从而有

$$0\in\partial f(\bar{x})+N_s(\bar{x})$$

即 $\bar{x}$ 是优化问题的一个关键点。

综上可知，网络模型的每一个聚点都是优化问题（1）的关键点，证毕。

3 实验仿真

神经网络最优化模式与常规的最优化算法的最大不同为神经网络模型的并行计算能力。得益于这种并行性能，神经网络模型已经被应用于广泛变化的科学工程应用中。为了检验微分包含的神经网络模型(2)的有效性，使用 MATLAB2012a 来仿真最优化问题式(1)，验证解轨迹 $x(t)$ 最终收敛于 D^* 中的一个最优解。由于 MATLAB2012a 的默认数据类型为 short，数据精度为保留小数点后四位，因此本文仿真的所有数据结果均保留四位小数。

例 1 (目标函数非光滑，约束函数光滑)

$$\min \quad f(\boldsymbol{x})=|x_1|+\cos(x_3)+2x_2^2$$
$$\text{s.t.} \quad \begin{aligned} & x_1^2+x_3^2-10\le 0 \\ & 2x_2+x_2^2+x_3-5\le 0 \\ & x_1+x_2+x_3-1=0 \end{aligned}$$

其中，目标函数为非光滑非凸，约束为光滑凸函数。取 5 组随机初始点 $(-1.5,1.5,1.5)^T$、$(-3,2,3)^T$、$(0.5,-0.5,2.5)^T$、$(-2,3,1)^T$、$(-1,1,2)^T$，其中，$x=(x_1,x_2,x_3)^T$ 的轨迹如图 1 所示，所有解轨迹均可以收敛到唯一的平衡点。其解为 $x=(-1.7921,-0.2156,3.0038)^T$。

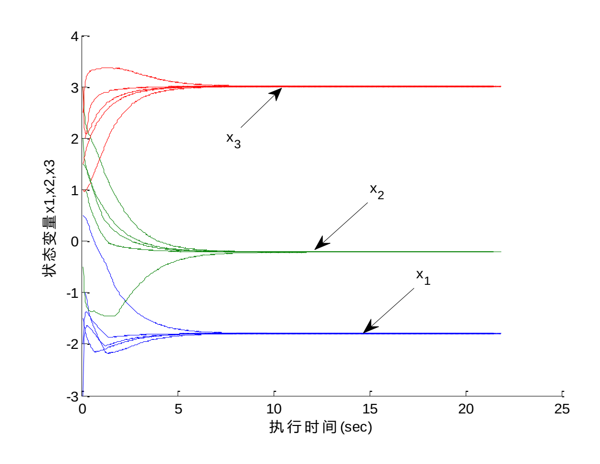


图 1 例 1 解收敛轨迹

例 2 (目标函数和约束函数均非光滑凸函数)

$$\text{Min} \quad f(\boldsymbol{x})=|x_1|+x_2^2\text{-}\cos(x^3)$$
$$\text{s.t.} \quad \begin{aligned} & x_1^2+x_2^2+x_3^2-8\le 0 \\ & |x_1|+x_2+x_3-10\le 0 \\ & x_1+x_2+x_3-4=0 \end{aligned}$$

在例 2 中，目标函数是非光滑凸函数，其实一个不等式约束为非光滑凸函数。选取五组初始点：$(-1,0,3)^T$、$(-2,-1,2)^T$、$(-3,2.5,0.5)^T$、$(-4,1,1.5)^T$、$(-1.5,1.5,1)^T$。状态向量 $x(t)$ 的解轨迹如图 2 所示，$x(t)$ 均可以收敛于唯一解。其中，解为 $x=(0.0043,0.2640,3.7377)^T$。

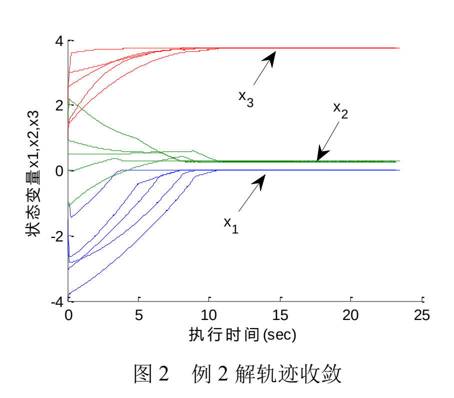


图 2　例 2 解轨迹收敛

4　结束语

本文提出了一种新的解决非光滑非凸优化的等式和不等式约束的递归神经网络模型。跟传统神经网络模型相比，本文所提出的递归神经网络模型不需要任何的惩罚函数，解决了惩罚参数在实际中难以计算的问题，同时，网络收敛速度得到了极大的提高。最后，本文使用了两个例子验证了本文所提递归神经网络求解非光滑凸优化问题的理论的正确性。接下来的工作，我们将探索其他神经网络模型(尤其是混沌神经网络)对非光滑非凸最优化问题进行求解。

参考文献：

[1] Tank D W, Hopfield J J. Simple neural optimization networks: an A//D converter, signal decision circuit, and a linear programming circuit [J]. IEEE Trans on Circuits Systems, 1986, 33 (5): 533–541.

[2] Kennedy M P, Chua L O. Neural networks for nonlinear programming [J]. IEEE Trans on Circuits Systems, 1988, 35 (5): 554–562.

[3] Zhang Shengwei, Constantinides A G. Lagrange programming neural networks [J]. IEEE Trans on Circuits and Systems Ⅱ: Analog and Digital Signal Processing, 1992, 39 (7): 441-452.

[4] Feng Ruibin, Leung ChiSing, Constantinides A G, *et al.* Lagrange programming neural network for nondifferentiable optimization problems in sparse approximation [J]. IEEE Trans on Neural Networks and Learning Systems, 2017, 28 (10): 2395-2407.

[5] Liang Junli, Leung Chising, So Hingcheung. Lagrange programming neural network approach for target localization in distributed MIMO radar [J]. IEEE Trans on Signal Processing, 2016, 64 (6): 1574-1585.

[6] Liang Junli, So Hingcheung, Leung Chising, *et al.* Waveform design with unit modulus and spectral shape constraints via lagrange programming neural network [J]. IEEE Journal of Selected Topics In Signal Processing, 2015, 9 (8): 1377-1386.

[7] Forti M, Nistri P, Quincampoix M. Generalized neural network for nonsmooth nonlinear programming problems [J]. IEEE Trans on Circuits and Systems I: Regular Papers, 2004, 51 (9): 1741–1754.

[8] Bian Wei, Xue XiaoPing. Subgradient-based neural networks for nonsmooth nonconvex optimization problems [J]. IEEE Trans on Neural Networks, 2009, 20 (6): 1024-1038.

[9] Bian Wei, Xue XiaoPing. Subgradient-based neural networks for nonsmooth convex optimization problems [J] IEEE Trans on Circuits and Systems I: Regular Papers, 2008, 55 (8): 2378-2391.

[10] Hu XiaoLin, Wang Jun. Solving pseudomonotone variational inequalities and pseudoconvex optimization problems using the projection neural network [J]. IEEE Trans on Neural Networks, 2006, 17 (6): 1487-1499.

[11] Cheng Long, Hou Zengguang, Lin Yingzi, *et al.* Recurrent neural network for non-smooth convex optimization problems with application to the identification of genetic regulatory networks [J]. IEEE Trans on Neural Networks, 2011, 22 (5): 714-726.

[12] Qin Sitian, Feng Jiqiang, Song JiaHui, *et al.* A one-layer recurrent neural network for constrained complex-variable convex optimization [J]. IEEE Trans on Neural Networks and Learning Systems, 2018, 29 (3): 534-544.

[13] Liu Xiaolan, Zhou Mi. A one-layer recurrent neural network for non-smooth convex optimization subject to linear inequality constraints [J]. Chaos, Solitons and Fractals, 2016 (87): 39-46.

[14] Guo Zhishan, Liu Qingshan, Wang Jun. A one-layer recurrent neural network for pseudoconvex optimization subject to linear equality constraints [J]. IEEE Trans on Neural Networks, 2011, 22 (12): 1982-1900.

[15] Liu Qingshan, Guo Zhishan, Wang Jun. A one-layer recurrent neural network for constrained pseudoconvex optimization and its application for dynamic portfolio optimization [J]. Neural Networks, 2012, 21 (1): 99-109.

[16] Liu Qingshan, Cao Jinde, Chen Guangrong. A novel recurrent neural network with finite-time convergence for linear programming [J]. Neural Computation, 2010, 22 (11): 2962-2978.

[17] Huang Yuancan, Yu Chuang. Improved lagrange nonlinear programming neural networks for inequality constraints [C]// Proc of International Joint Conference on Neural Networks. 2007: 962-966

[18] 喻昕，许治健，陈昭蓉，等. 拉格朗日神经网络解决带等式和不等式约束的非光滑非凸优化问题 [J]. 电子与信息学报，2017, 39 (8): 1950-1955. (Yu Xin, Xu Zhijian, Zhen Zhaorong, *et al.* Lagrange neural network for nonsmooth nonconvex optimization problems with equality and inequality constrains [J]. Journal of Electronics & Information Technology, 2017, 39 (8): 1950-1955.)

[19] Aubin J P, Cellina A. Differential inclusion: set-valued maps and viability theory [M]. Berlin: Springer-Verlag, 1984.

[20] Clarke F H. Optimization and nonsmooth analysis [M]. New York, USA: Wiley, 1983.

[21] 喻昕，李晨宇，许治健，等. 一种新型拉格朗日神经网络解决非光滑优化问题 [J]. 计算机应用研究，2016, 33 (11): 3261-3264, 3269. (Yn Xin, Li Chenyu, Xu Zhijian, *et al.* Novel lagrange neural network for nonsmooth optimization problems [J]. Application Research of Computers, 2016, 33 (11): 3261-3264, 3269.)

[22] 喻昕，于琰，谢缅. 光滑拉格朗日神经网络解决非光滑最优化问题 [J]. 计算机应用研究，2014, 31 (5): 1349-1352. (Yu Xin, Yu Yan, Xie Mian. Smoothing Lagrange neural network for constrained nonsmooth

optimization problems [J]. Application Research of Computers, 2014, 31 (5): 1349-1352.)

[23] 喻昕，谢缅，李晨宇. 光滑神经网络解决非李普西茨优化问题的研究 [J]. 计算机工程与科学, 2015, 37 (12): 2262-2269. (Yu Xin, Xie Mian, Li Chenyu. Solving non-Lipschitz optimization problems by smoothing neural networks [J]. Computer Engineering and Science, 2015, 37 (12): 2262-2269.)